



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) EP 1 030 254 A1

(12) EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:  
23.08.2000 Patentblatt 2000/34

(51) IntCl.7: G06F 17/30

(21) Anmeldenummer: 99103194.9

(22) Anmeldetag: 18.02.1999

(84) Benannte Vertragsstaaten:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Benannte Erstreckungsstaaten:  
AL LT LV MK RO SI

- Scholz, Gero Dr.  
61194 Niddatal Assenheim (DE)
- Hesse, Andreas  
85521 Ottobrunn (DE)
- Bröseler, Peter Dr.  
81829 München (DE)

(71) Anmelder: ed&m,  
Software design & management GbmH & Co. KG  
81737 München (DE)

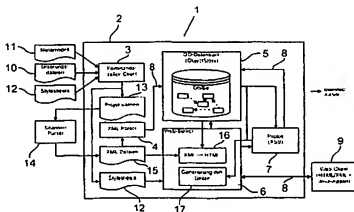
(74) Vertreter: Manitz, Finsterwald & Partner  
Postfach 22 16 11  
80506 München (DE)

(72) Erfinder:  
• Schulz, Helge  
81739 München (DE)

(54) Verfahren und System zum Verwalten von Dokumenten

(57) Es wird ein Verfahren für ein System zum Er-schließen und Verwalten von in elektronischer Form vorliegenden Dokumenten beschrieben. Bei dem Ver-fahren wird der Quellinhalt von in einem Quelldaten-format oder in mehreren Quelldatenformaten vorliegenden Dokumenten eines ersten Typs automatisch analysiert und in ein einheitliches Zieldatenformat konvertiert. Dabei werden zumindest für einen Teil des Quellinhalts der Dokumente ersten Typs gemäß für das Quelldaten-format bzw. für die Quelldatenformate kennzeichnenden Beschreibungsregeln automatisch Kennzeichnungen generiert und diese zusammen mit zumindest einem Teil des Quellinhalts in dem Zieldatenformat als Doku-

mente zweiten Typs abgespeichert. Zusätzlich werden diese Kennzeichnungen in einer Datenbank abgespei-chert und der in den Dokumenten zweiten Typs gespei-cherter Quellinhalt über ein Hypertextsystem angezeigt, wobei zumindest ein Teil des Quellinhalts, für den Kenn-zeichnungen generiert wurden, als Hyperlinks ange-zeigt wird, die Information zur Dereferenzierung der Hy-perlinks aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet wird und zusätzlich über logikbasierte Anfrageelemente automatisch Dokumen-te eines dritten Typs erzeugt werden können, die über das Hypertextsystem angezeigt werden und Hyperlinks enthalten, die aus den in der Datenbank abgespeicher-ten Kennzeichnungen abgeleitet werden.



EP 1 030 254 A1

Printed by Jouve, 75001 PARIS (FR)

received  
SR 0101 4/10/00  
CIC 000001

## Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Verfahren sowie ein System zum Erschließen und Verwalten von in elektronischer Form vorliegenden Dokumenten.

[0002] Seit der Einführung des Internets hat sich für die Verwaltung von Dokumenten in zunehmendem Maße das Hypertext Transfer Protokoll (HTTP) unter Verwendung des Hypertext Markup Language (HTML)-Formate durchgesetzt. Dabei erfolgt eine entsprechende Dokumentenverwaltung nicht nur im Rahmen des Internets, sondern auch in lokalen Anwendungen, wie beispielsweise auf der HTML-Sprache basierenden Hilfesystemen.

[0003] Kennzeichnend für diese Systeme ist es, daß von bestimmten, durch sogenannte Hyperlinks gekennzeichneten Bereichen der Dokumente Verweise zu Stellen in anderen oder auch denselben Dokumenten existieren, über die, beispielsweise durch Anklicken mit der Maus, eine Navigation zwischen unterschiedlichen Dokumenten und/oder darin befindlichen Bezugspunkten möglich ist.

[0004] Die Verweise sind dabei üblicherweise fests als relative oder absolute Pfade in den jeweiligen Dokumenten angegeben. Die durch diese Verweise definierte Struktur ist daher starr und unflexibel, da bei einer Änderung der Positionen der einzelnen Dokumente innerhalb des Dateisystems und/oder Verschiebung von Bezugspunkten innerhalb oder zwischen Dokumenten die Verweise jeweils geändert werden müssen.

[0005] Weiterhin ist lediglich eine einfache Punkt zu Punkt Navigation zwischen den Dokumenten möglich, eine weitergehende, komplexere Navigation, die bestimmte Eigenschaften von in den Dokumenten enthaltenen Informationsobjekten berücksichtigt, ist mit diesem System nicht möglich.

[0006] Es ist eine Aufgabe der vorliegenden Erfindung, ein Verfahren und ein System zum Erschließen und Verwalten von in elektronischer Form vorliegenden Dokumenten anzugeben, das flexibler als die bekannten Verfahren und Systeme einsetzbar ist.

[0007] Diese Aufgabe wird durch ein Verfahren gelöst, bei dem der Quellinhalt von in einem Quelldatenformat oder in mehreren Quelldatenformaten vorliegenden Dokumenten eines ersten Typs automatisch analysiert und in ein einheitliches Zieldatenformat konvertiert wird, wobei zumindest für einen Teil des Quellinhalts der Dokumente ersten Typs gemäß für das Quelldatenformat bzw. für die Quelldatenformate kennzeichnenden Beschreibungsregeln automatisch Kennzeichnungen generiert und diese zusammen mit zumindest einem Teil des Quellinhalts in dem Zieldatenformat als Dokumente zweiten Typs abgespeichert werden, zusätzlich diese Kennzeichnungen in einer Datenbank abgespeichert werden, und in den Dokumenten zweiten Typs gespeicherte Quellinhalt über ein Hypertextsystem angezeigt wird, wobei zumindest ein Teil des Quellinhalts, für den Kennzeichnungen generiert wurden, als Hyperlinks angezeigt wird, die Information zur Dereferenzierung der Hyperlinks aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet wird und zusätzlich über logikbasierte Anfrageelemente automatisch Dokumente eines dritten Typs erzeugt werden können, die über das Hypertextsystem angezeigt werden und Hyperlinks enthalten, die aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet werden.

[0008] Ein erfindungsgemäßes System umfaßt eine Client-Server-Struktur, wobei der Server folgende Elemente umfaßt.

- eine Schnittstelle zum Übernehmen von in einem Quelldatenformat oder in mehreren Quelldatenformaten vorliegenden Dokumenten eines ersten Typs von dem Client,
- eine Analyse- und Konvertiereinheit zum automatischen Analysieren des Quellinhalts der Dokumente ersten Typs und zum Konvertieren des Quellinhalts in ein einheitliches Zieldatenformat, wobei durch die Analyse- und Konvertiereinheit zumindest für einen Teil des Quellinhalts der Dokumente ersten Typs gemäß für das Quelldatenformat kennzeichnenden Beschreibungsregeln automatisch Kennzeichnungen generierbar und diese zusammen mit zumindest einem Teil des Quellinhalts in dem Zieldatenformat als Dokumente zweiten Typs abgespeicherbar sind,
- eine Datenbank und eine Extrahiereinheit zum Abspeichern dieser Kennzeichnungen in der Datenbank,
- und ein Hypertextsystem zum Anzeigen des in den Dokumenten zweiten Typs gespeicherten Quellinhalts, wobei zumindest ein Teil des Quellinhalts, für den Kennzeichnungen generiert wurden, als Hyperlinks anzeigbar sind, die Information zur Dereferenzierung der Hyperlinks aus den in der Datenbank abgespeicherten Kennzeichnungen ableitbar sind und zusätzlich über logikbasierte Anfrageelemente automatisch Dokumente eines dritten Typs erzeugbar sind, die über das Hypertextsystem anzeigbar sind und Hyperlinks enthalten, die aus den in der Datenbank abgespeicherten Kennzeichnungen ableitbar sind.

[0009] Erfindungsgemäß werden somit in beliebigen Quelldatenformaten vorliegende Dokumente, die in ihrer Gesamtheit als Dokumente eines ersten Typs bezeichnet werden, automatisch analysiert und in ein einheitliches Zieldatenformat konvertiert. Dieses einheitliche Zieldatenformat enthält dabei Kennzeichnungen für die in den Dokumenten ersten Typs enthaltenen Informationen und Beziehungen, durch die diese maschinell identifiziert und als Datenobjekte in eine Datenbank übernommen werden können. Diese Kennzeichnungen werden automatisch ebenfalls gemäß vorgegebenen Beschreibungsregeln erzeugt, die ebenfalls in Dokumenten abgespeichert sind und im folgenden auch als Metamodell

bezeichnet werden.

[0010] Dieses Metamodell muß je nach gewünschten Quelldatenformaten einmalig definiert werden, so daß durch das Metamodell alle für die jeweiligen Dokumente relevanten Objekte und Beziehungen beschreibbar sind.

[0011] Das Metamodell basiert bevorzugt auf einer festen Elementspezifikation, die im weiteren auch als Meta-Metamodell bezeichnet wird, und ist basierend auf diesem Meta-Metamodell frei definierbar. Dadurch ist gewährleistet, daß innerhalb des Meta-Metamodells das Metamodell in beliebiger Weise definiert werden kann und eine Vielzahl von unterschiedlichen Dokument-, Objekt- und Beziehungstypen definiert und verwaltet werden können.

[0012] Wesentlich für die vorliegende Erfindung ist es, daß durch das definierte Metamodell die in den Dokumenten zweiten Typs generierten Kennzeichnungen automatisch extrahiert werden können und in einer Datenbank abgespeichert werden. Es werden somit nicht die Dokumente selbst und ihr struktureller Aufbau als Daten gespeichert, sondern die Meta-Daten als die in den Dokumenten beschriebenen Sachverhalte und die Beziehungen zwischen diesen Sachverhalten, die typisiert gemäß dem hinterlegten Metamodell gespeichert werden.

[0013] Der Zugriff auf die in den Dokumenten gespeicherten Informationen erfolgt dann über in den Dokumenten enthaltene Hyperlinks, wobei die Information zur Dereferenzierung der Hyperlinks aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet wird. Die Navigation innerhalb der Dokumente erfolgt somit über die Dokumente selbst, wobei von den Stielen eines Dokuments, an dem Objekte bzw. Beziehungen beschrieben sind, diese Beziehungen als Hyperlinks unter Verwendung der in der Datenbank abgespeicherten Meta-Daten verfolgt werden können.

[0014] Mit dem erfindungsgemäßen Verfahren und dem erfindungsgemäßen System kann beispielsweise bei einer Mehrdeutigkeit eines über einen Hyperlink erreichbaren Zieles eine Auswahl von möglichen Zielen aufgebündelt werden, aus der das gewünschte Ziel auswählbar ist. Dabei ist es zusätzlich möglich, daß über logikbasierte Anfrageelemente automatisch Dokumente eines dritten Typs erzeugt werden, die über das Hypertextsystem angezeigt werden und wiederum Hyperlinks enthalten, die aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet sind. Durch diese logikbasierte Anfrageelemente wird praktisch eine unbegrenzte Flexibilität bei der Formulierung möglicher Anfragen erreicht, die auf der Grundlage des frei konfigurierbaren Metamodells basiert.

[0015] Wahlweise können diese logikbasierten Anfragen dabei als Standardfragen vordefiniert sein oder auch als individuelle Anfragen in einer vorgegebenen Anfragesprache formuliert werden, um auf diese Weise eine noch größere Flexibilität zu erreichen.

[0016] Bevorzugt sind in der festen Elementspezifikation Deklarationen von Objekttypen, von Dokumenttypen und von Beziehungstypen definiert. Dieses feste Meta-Metamodell ist einerseits relativ einfach gehalten und andererseits trotzdem mächtig genug, eine Vielzahl von Sachverhalten beschreiben zu können.

[0017] Nach einer weiteren vorteilhaften Ausführungsform der Erfindung wird als Zieldatenformat das XML-(Extensible Markup Language)-Format verwendet. Bei dem XML-Format handelt es sich um ein standardisiertes Format, mit dem sehr flexibel Eigenschaften von Objekten und Beziehungen beschrieben werden können. Weiterhin ist es möglich, die in dem XML-Format vorliegenden Zieldokumente in das HTML (Hypertext Markup Language)-Format zu konvertieren, so daß die Navigation zwischen den Dokumenten mittels eines HTML-Browsers möglich ist und die HTML-Dokumente über einen Web-Server verwaltet werden können. Bei der Verwendung von XML-fähigen Browsern kann die Konvertierung nach HTML selbstverständlich entfallen.

[0018] Als Anfragesprache wird bevorzugt Prolog verwendet, das deterministische Antworten liefert und keine doppelten Ergebnissätze zurückgibt. Weiterhin wird als Datenbank vorteilhaft eine objektorientierte Datenbank verwendet, mit der Prolog eng gekoppelt zusammenarbeitet, so daß Objekte und Beziehungen nicht doppelt gespeichert werden müssen. Bei der Formulierung von individuellen Anfragen kann dabei auf Prolog-Prädikate zurückgegriffen werden, die direkt den Begriffen des Metamodells entsprechen.

[0019] Weitere vorteilhafte Ausführungsformen der Erfindung sind in den Unteransprüchen angegeben.

[0020] Die Erfindung wird nachfolgend anhand eines Ausführungsbeispiels unter Bezugnahme auf die einzige Figur näher beschrieben.

[0021] In der Figur ist das erfindungsgemäße Client/Server-System 1 in vereinfachter Form in einem Blockdiagramm dargestellt.

[0022] Das Client/Server-System 1 umfaßt eine Server-Instanz 2, die auf einer Server-Maschine läuft und sich von Clients unter einer bestimmten Netzwerkadresse ansprechen läßt. Dabei können eine Vielzahl von Server-Instanzen 2 gleichzeitig unter verschiedenen Netzwerkadressen unabhängig voneinander auf einem Rechner arbeiten.

[0023] Die Server-Instanz 2 umfaßt einen Kommandozeilen-Client 3, einen XML-Parser 4, eine objektorientierte Datenbank 5, einen erweiterten Web-Server 6 sowie ein Prolog-System 7, die untereinander Daten austauschen bzw. sich gegenseitig aufrufen, wie es durch eine Vielzahl von Pfeilen 8 dargestellt ist.

[0024] Weiterhin ist ein Web-Client 9 abgebildet, der auf einem Client-Rechner läuft und mit der Server-Instanz 2 ebenfalls Daten austauscht.

[0025] Mit 10 bezeichnete Ursprungsdateien bilden in einem oder in mehreren unterschiedlichen Quelldatenformaten vorliegende Dokumente eines ersten Typs, deren Inhalt in das erfindungsgemäße Client/Server-System 1 eingegeben

wird. Beispiele für mögliche Dokumenttypen sind Computerprogramme, Personalstammbblätter einer Personalverwaltung, Wörterbücher, Lexika oder dergleichen.

[0026] Für die Gesamtheit dieser Dokumenttypen existiert ein Metamodell 11, durch das die in den jeweiligen Dokumenttypen enthaltenen Informationen und Beziehungen beschrieben werden können.

[0027] Zusätzlich können je Dokumenttyp sogenannte Stylesheets 12 vorgesehen sein, durch die ein Anzeigelayout der Inhalte der Ursprungsdateien 10 definiert wird.

[0028] Die Ursprungsdateien 10, das Metamodell 11 und die Stylesheets 12 gelangen durch den Aufruf des Kommandozeilen-Clients 3 in die Server-Instanz 2. Dazu wird der Kommandozeilen-Client 3 in den jeweiligen Client-Rechner geladen und überträgt von dort die Ursprungsdateien 10, das Metamodell 11 und die Stylesheets 12 über das Netzwerk zum Server-Rechner, wo sie in einer Projektdokumentenablage als Projektdaten 13 abgelegt werden.

[0029] Damit die Inhalte der Ursprungsdateien 10 von der Server-Instanz 2 automatisch analysiert und erfaßt werden können, muß für jedes der verwendeten Quelldatenformate ein geeigneter Scanner/Parser 14 vorgesehen sein, durch den gemäß dem Metamodell 11 die Ursprungsdateien 10 in Dokumento eines zweiten Typs mit einem einheitlichen Zieldatenformat konvertiert werden können. Dieses Zieldatenformat kann beispielsweise das sogenannte XML-Format sein, so daß die in der Figur dargestellten XML-Dateien 15 erzeugt werden.

[0030] Bei dieser Konvertierung werden die in den Ursprungsdateien 10 enthaltenen Inhalte sowie die Beziehungen zwischen diesen Inhalten von dem Scanner/Parser 14 gemäß dem Metamodell 11 und gemäß dem XML-Nutzungskonzept markiert, so daß die abgespeicherten XML-Dateien 15 neben den ursprünglichen Inhalten der Ursprungsdateien 10 Kennzeichnungen dieser Inhalte enthalten.

[0031] Die auf diese Weise in einem einheitlichen Zieldatenformat vorliegenden XML-Dateien 15 werden durch den XML-Parser 4 untersucht, der anhand des Metamodells die automatisch erzeugten Kennzeichnungen aus den XML-Dateien 15 extrahiert und in die objektorientierte Datenbank 5 einträgt. Es werden somit nicht die Dokumenteninhalte selbst und ihr struktureller Aufbau in der Datenbank gespeichert, sondern ausschließlich die automatisch generierten Kennzeichnungen, welche die in den Dokumenten beschriebenen Sachverhalte und die Beziehungen zwischen diesen Sachverhalten gemäß dem hinterlegten Metamodell beschreiben.

[0032] Der gesamte Import der Ursprungsdateien 10 von seiner Übertragung auf den Server-Rechner bis zum Abspeichern in der Datenbank 5 ist unteilbar und erfolgt in einer Transaktion. Hat diese Transaktion Erfolg, so werden die XML-Dateien 15 in einer XML-Dateiablage für die Hyperextnavigation freigegeben. Schlägt die Transaktion fehl, so werden, falls vorhanden, die alte Projektdatei wiederhergestellt, die Datenbankänderung zurückgenommen und weiterhin die bisherigen XML-Dateien benutzt.

[0033] Liegen die Ursprungsdateien 10 bereits in dem Zieldatenformat, d.h. im Beispiel im XML-Format oder einem XML-konformen Format (beispielsweise HTML) vor, so werden die Inhalte der Ursprungsdateien von dem Parser 14 transparent weitergereicht bzw. lediglich eine Markierung der Objekte und Beziehungen durchgeführt. Enthalten Dokumente Graphiken oder ähnliche Elemente, so kann der Parser 14 in diesem Schritt die Ursprungsdateien 10 in mehrere Dateien zerlegen.

[0034] Die erzeugten XML-Dateien 15 werden dem erweiterten Web-Server 6 zugeführt, der Teil der Server-Instanz 2 ist. Der Web-Server 6 stellt die entsprechenden Hypertext-Navigationsmöglichkeiten bereit, so daß die XML-Dateien 15 über einen Hypertext-Browser angezeigt werden können. Handelt es sich bei dem Hypertext-Browser um einen HTML-Browser so wandelt der Web-Server 6 die in XML-Format vorliegenden Dateien beim Versenden an den Web-Client 9 online nach HTML um, wie es durch einen XML→HTML-Konverter 16 angedeutet ist. Der XML→HTML-Konverter 16 wandelt dabei nur die Kennzeichnungen in das HTML-Format um, die Objekte und Beziehungen laut des jeweiligen Metamodells 11 markieren. Dadurch wird ermöglicht, daß die Ursprungsdateien 10 im HTML-Format korrekt angezeigt werden. Weiterhin wird bei der Umwandlung eine besondere Markierung des Zielobjekts bei der Navigation erzeugt, so daß dieses bei der Anzeige über den Web-Client 9 hervorgehoben und vom Benutzer leichter auffindbar ist. Die Art der Hervorhebung wird dabei durch die Stylesheets 12 bestimmt, auf die die Dokumente verweisen und welche automatisch von dem Web-Client 9 abgerufen werden.

[0035] Der Web-Client 9 kann neben den Dokumenten auch Programmcode beispielsweise in Form von Java und/oder Javascript, enthalten, der Menüs und Dialoge realisiert, die zur Ausführung von Suchen sowie zur Delimitation und Ausführungen von Anfragen und dergleichen dienen.

[0036] Neben der Navigation innerhalb der Dokumente über die aus den in der Datenbank 5 gespeicherten Metadaten abgeleiteten Hyperlinks können logikorientierte Anfragen durch das Prolog-System 7 abgearbeitet werden. Dabei können die Anfragen in Form von Standardanfragen vorliegen, es kann bei der Formulierung der Anfragen aber auch auf Prolog-Prädikate zurückgegriffen werden, die direkt den Begriffen aus dem verwendeten Metamodell 11 entsprechen. Das Ergebnis einer solchen Anfrage ist dann ein Dokument 17 mit Verweisen auf Dokumentsteilen, an denen die zum Anfrageergebnis gehörenden Objekte bzw. Beziehungen definiert sind.

[0037] Erste Ergebnisse können dabei im Hypertext-Browser schon vor dem Ende der gesamten Anfragen angezeigt werden, da das Prolog-System 7 diese satzweise übergibt. Die Transaktionssteuerung erfolgt so, daß die Anfrage sich immer auf den Stand der Datenbank 5 bezieht, wie er beim Start der Anfrage vorlag.

[0038] Wie bereits erwähnt wurde, ist es für das erfindungsgemäße Verfahren und System wesentlich, daß zum einen ein festes Meta-Metamodell existiert, auf dessen Basis ein frei konfigurierbares Metamodell zur Typisierung der Dokumenteninhalte definierbar ist. Dabei wird dieses Metamodell für die Gesamtheit der zu verwendenden Quelldatenformate entsprechend der in den Dokumenten gespeicherten, interessierenden Objekte und Beziehungen einmalig definiert und der automatischen Extraktion der Objekte und Beziehungen zugrundegelegt.

[0039] Eine mögliche Metamodellspezifikation besteht beispielsweise aus den folgenden drei Typen von Einträgen:

1. Deklaration von Dokumenttypen
2. Deklaration von Objekttypen
3. Deklaration von Beziehungstypen.

[0040] Die Deklaration von Dokumenttypen kann optionale Angaben von Dokumenttyp-Supertypen sowie die Information enthalten, ob Dokumente dieses Typs für die Volltextsuche indiziert werden. Diese Deklaration kann beispielsweise in der folgenden Notation erfolgen:

[0041] Dokumenttyp Dokumenttypname  
{ ist (ein | eine) Dokumenttypname }  
{ ist nicht volltextindizierbar }

[0042] Dabei sind Alternativen "a oder b" dargestellt als ( a | b ). Teile die wiederholt werden können sind mit {} gekennzeichnet, optionale Teile sind mit [] geklammert.

[0043] Die Deklaration von Objekttypen enthält optionale Angaben der Dokumenttypen, in denen sie definiert werden können, sowie die optionale Angaben von Objekttyp-Supertypen in der Form

[0044] Objekttyp Objekttypname  
{ ist (ein | eine) Objekttypname }  
{ ist definiert in Dokumenttypname }

[0045] Die Deklaration von Beziehungstypen enthält die Angaben der Ziel-Resource, bzw. alle von diesen abgeleiteten Subtypen, und deren Kardinalitäten mit optionaler Angabe der Dokumenttypen, in denen sie definiert werden können und optionaler Angabe von Beziehungstyp-Supertypen. Diese kann formuliert werden als

[0046] Beziehungstyp Beziehungstypname [ alias Alias Name ]  
von n1 bis n2 1\* Ressourcetyppname  
nach m1 bis m2 1\* Ressourcetyppname  
( ist Teilmenge von Beziehungstypname )  
( ist definiert in Dokumenttypname )

[0047] Alle vorkommenden Typen müssen in genau einem Eintrag deklariert werden. Jeder Typ muß durch seinen Namen eindeutig gekennzeichnet sein, wobei Namen mit gleicher Buchstabenreihenfolge aber unterschiedlicher Groß-/Kleinschreibung als gleich gelten und für Prolog die Anfangsbuchstaben der Namen ggf. in Kleinbuchstaben umgewandelt werden.

[0048] Zur weiteren Verdeutlichung ist im folgenden ein Ausschnitt eines Metamodells dargestellt:

Objekttyp Anwendungstyp ist definiert in Quellcode	// Objekte des Typs "Anwendungstyp" // werden in Dokumenten // des Typs "Quellcode" definiert
ist ein Typ	// Vererbungsbeziehung

Beziehungstyp ändert	
von Fähigkeit nach Attribut kommt vor in Quellcode	// Quelltyp // Zieltyp // Beziehungen des Typs "ändert" // kommen in Dokumenten // des Typs "Quellcode" vor

[0049] Dieser Ausschnitt aus einer Metamodelldefinition sieht überführt in XML folgendermaßen aus, wobei die Teile, die unverändert aus der Eingabe übernommen wurden, fett dargestellt sind.

[0050] <Objekt ID="Anwendungstyp" Rolle = "Definition">Objekttyp

[0051] Anwendungstyp

Ist definiert in

<Dokument IDREF = "Quellcode" Rolle = "Inhalt">

Quellcode </Dokument>

Ist ein <Objekt IDREF = "Typ" Rolle = "Supertyp">Typ</Objekt> </Objekt>

[0053] <Beziehung ID="aendert" Rolle="Definition">Beziehungstyp aendert  
von <Objekt IDREF="Faehigkeit" Rolle="Quelle">Faehigkeit </Objekt>  
nach <Objekt IDREF="Attribut" Rolle="Ziel">Attribut</Objekt>  
kommt vor in <Dokument IDREF="Quellcode" Rolle="Inhalt">

Quellcode</ Dokument>

< Beziehung>

[0053] Die automatisch eingefügten Kennzeichnungen (in spitzen Klammern) dienen dazu, die im Dokument enthaltenen und für einen menschlichen Leser verstehbare Informationen für die weitere Verarbeitung explizit zu kennzeichnen und damit maschinell extrahierbar zu machen. Erfindungsgemäß werden lediglich die in den spitzen Klammern angegebenen Meta-Daten in der objektorientierten Datenbank gespeichert, während der außerhalb der spitzen Klammern vorhandene, eigentliche Dokumentinhalt nur für den menschlichen Leser gedacht ist. Die Dokumente werden in dem Dateisystem des Servers geordnet verwaltet und auf Anforderung an den Hypertext-Browser zur Anzeige übermittelt.

[0054] Als Anfragesprache dienen erfindungsgemäß Prolog-Ausdrücke mit der Semantik von erweitertem Datalog. Die Anfragen werden von dem Prolog-System 7 bearbeitet, das eng mit der objektorientierten Datenbank 5 gekoppelt ist, wobei die Anfragen die folgende Form besitzen:

Anfrage	::= Ausdruck <sup>1</sup>
Ausdruck	::= Term   '(' Ausdruck ')'   Ausdruck '(' ':' '!' ')' Ausdruck
Term	::= Prädikat '(' 'Argument' '(' 'Argument' ')' )'
Argument	::= Variable   Konstante   '_'
Variable	::= 'A' - 'Z'   'a' - 'z'   'A' - 'Z'   '0' - '9' )'
Konstante	::= String   Zahl   Prädikat
String	::= "'Zeichen' "

[0055] Dabei besitzt ein Semikolon die Bedeutung einer Disjunktion während das Komma die Bedeutung einer Konjunktion besitzt.

In einem Anfrageterm kann der Benutzer auf Prädikate zugreifen, die von dem Prolog-System selbst, durch selbst definierte Regeln, durch das Meta-Metamodell oder durch das Metamodell definiert sind. Prädikate sind spezielle Funktionen, deren Ergebnismenge nur aus den Wahrheitswerten "wahr" und "falsch" besteht, wobei sich als Argumente Konstanten oder sogenannte freie Variablen einsetzen lassen. Dabei ist auch die Verwendung einer universellen freien Variable als Platzhalter für uninteressante Dinge möglich (dargestellt durch '\_').

[0056] Bei der Auswertung der Anfrage sucht das Prolog-System 7 alle möglichen Belegungen der freien Variablen heraus, bei denen der Anfrageausdruck insgesamt den Wahrheitswert "wahr" ergibt. Das Ergebnis besteht aus einer Liste der möglichen Belegungen, wobei jede Belegung in einer eigenen Zeile steht.

[0057] Handelt es sich bei den Werten um durch das Metamodell definierte Ressourcen (Dokumente oder Objekte) so wird der Typ- und Ressourcenname als Hyperlink ausgegeben. Bei Beziehungen wird jeweils sowohl der Wert für Ziel als auch für Quelle als Hyperlink ausgegeben. Die Hyperlinks von einem Ressourcennamen verweisen dabei wie üblich auf die Definitionsstelle des Objekts bzw. auf das Dokument.

[0058] Beispiele für Anfragen sind zur Verdeutlichung im folgenden aufgeführt:

- Welche Dokumente verwaltet die Server-Instanz? Anfrage: dokument (X).
- Welche Objekte vom Typ "C-Funktion" gibt es? Anfrage in Prolog: object (X, "C-Funktion", \_). Anfrage als vordefiniertes Prädikat: c-Funktion (X, \_).
- Welche Beziehungen vom Typ "ruft" gibt es? Anfrage in Prolog: relation (X, "ruft", \_, \_). Anfrage als vordefiniertes Prädikat: ruft (X, \_, \_).

Bezugszeichenliste

[0059]

- 5    1    Client/Server-System
- 2    Server-Instanz
- 3    Kommandozeilen-Client
- 4    XML-Parser
- 5    objektorientierte Datenbank
- 10   6    Web-Server
- 7    Prolog-System
- 8    Pfeile
- 9    Web-Client
- 10   Ursprungsdateien
- 15   11   Metamodell
- 12   Stylesheet
- 13   Projektdateien
- 14   Scanner/Parser
- 15   XML-Dateien
- 20   16   XML-→HTML-Converter
- 17   temporäre Dateien

Patentansprüche

- 25   1.    Verfahren zum Erschließen und Verwalten von in elektronischer Form vorliegenden Dokumenten, bei dem der  
     Quellinhalt von in einem Quelldatenformat oder in mehreren Quelldatenformaten vorliegenden Dokumenten eines  
     ersten Typs automatisch analysiert und in ein einheitliches Zieldatenformat konvertiert wird, wobei zumindest für  
30    einen Teil des Quellinhalts der Dokumente ersten Typs gemäß für das Quelldatenformat bzw. für die Quelldaten-  
     formate kennzeichnenden Beschreibungsregeln automatisch Kennzeichnungen generiert und diese zusammen  
     mit zumindest einem Teil des Quellinhalts in dem Zieldatenformat als Dokumente zweiten Typs abgespeichert  
     werden, zusätzlich diese Kennzeichnungen in einer Datenbank abgespeichert werden, und der in den Dokumenten  
     zweiten Typs gespeicherte Quellinhalt über ein Hypertextsystem angezeigt wird, wobei zumindest ein Teil des  
35    Quellinhalts, für den Kennzeichnungen generiert wurden, als Hyperlinks angezeigt wird, die Information zur De-  
     referenzierung der Hyperlinks aus den in der Datenbank abgespeicherten Kennzeichnungen abgeleitet wird und  
     zusätzlich über logikbasierte Anfragelocale automatisch Dokumente eines dritten Typs erzeugt werden können,  
     die über das Hypertextsystem angezeigt werden und Hyperlinks enthalten, die aus den in der Datenbank abge-  
     speicherten Kennzeichnungen abgeleitet werden.
- 40   2.    Verfahren nach Anspruch 1,  
     dadurch gekennzeichnet,  
     daß eine feste Elementspezifikation vorgegeben ist und daß die Beschreibungsregeln basierend auf der festen  
     Elementspezifikation frei definierbar sind.
- 45   3.    Verfahren nach Anspruch 2,  
     dadurch gekennzeichnet,  
     daß in der festen Elementspezifikation Deklarationen von Objekttypen, von Dokumenttypen und von Beziehungs-  
     typen definiert sind.
- 50   4.    Verfahren nach Anspruch 1, 2 oder 3,  
     dadurch gekennzeichnet,  
     daß die Beschreibungsregeln in einem Dokument insbesondere ersten Typs abgespeichert sind.
- 55   5.    Verfahren nach einem der vorhergehenden Ansprüche,  
     dadurch gekennzeichnet,  
     daß als Zieldatenformat das XML-(Extensible Markup Language)-Format verwendet wird.
- 6.    Verfahren nach einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet,

daß die in dem Zieldatenformat vorliegenden Dokumente von einem Hypertext-Server verwaltet und in einem Hypertext-Browser angezeigt werden, insbesondere daß die in dem Zieldatenformat vorliegenden Dokumente in das HTML-(Hypertext Markup Language)-Format umgewandelt, die erzeugten HTML-Dokumente von einem Web-Server verwaltet und in einem Web-Browser angezeigt werden.

7. Verfahren nach einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet,

daß die logikbasierten Anfrageelemente als vorgegebene Standardanfragen definiert sind und/oder als individuelle Anfragen in einer vorgegebenen Anfragesprache formuliert werden.

8. Verfahren nach Anspruch 7,

dadurch gekennzeichnet,

daß als Anfragesprache Prolog verwendet wird.

9. Verfahren nach einem der vorhergehenden Ansprüche,

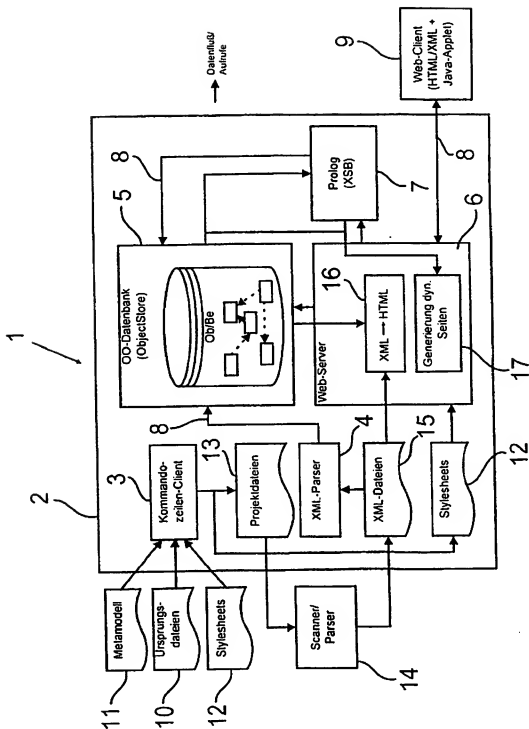
dadurch gekennzeichnet,

daß als Datenbank eine objektorientierte Datenbank verwendet wird.

10. System zum Erschließen und Verwalten von in elektronischer Form vorliegenden Dokumenten (10), insbesondere für ein Verfahren nach einem der vorhergehenden Ansprüche, mit einer Client-Server-Struktur, wobei der Server (2) umfaßt

- eine Schnittstelle (3) zum Übernehmen von in einem Quelldatenformat oder in mehreren Quelldatenformaten vorliegenden Dokumenten (10) eines ersten Typs von dem Client,
- eine Analyse- und Konvertiereinheit (14) zum automatischen Analysieren des Quellinhalts der Dokumente (10) ersten Typs und zum Konvertieren des Quellinhalts in ein einheitliches Zieldatenformat, wobei durch die Analyse- und Konvertiereinheit (14) zumindest für einen Teil des Quellinhalts der Dokumente (10) ersten Typs gemäß für das Quelldatenformat kennzeichnenden Beschreibungsregeln (11) automatisch Kennzeichnungen generierbar und diese zusammen mit zumindest einem Teil des Quellinhalts in dem Zieldatenformat als Dokumente (15) zweiten Typs abspeicherbar sind,
- eine Datenbank (5) und eine Extrahiereinheit (4) zum Abspeichern dieser Kennzeichnungen in der Datenbank (5),
- und ein Hypertextsystem (6, 9) zum Anzeigen des in den Dokumenten (15) zweiten Typs gespeicherten Quellinhalts, wobei zumindest ein Teil des Quellinhalts, für den Kennzeichnungen generiert wurden, als Hyperlinks anzeigbar sind, die Information zur Dereferenzierung der Hyperlinks aus den in der Datenbank (5) abgespeicherten Kennzeichnungen ableitbar sind und zusätzlich über logikbasierte Anfrageelemente automatisch Dokumente (17) eines dritten Typs erzeugbar sind, die über das Hypertextsystem anzeigbar sind und Hyperlinks enthalten, die aus den in der Datenbank abgespeicherten Kennzeichnungen ableitbar sind.







Europäisches  
Patentamt

# EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung  
EP 99 10 3194

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Bezieht Anspruch	KLASSIFIKATION DER ANMELDUNG
A	US 5 649 192 A (STUCKY ROBERT BRUCE) 15. Juli 1997 (1997-07-15) * Spalte 2, Zeile 31 - Spalte 6, Zeile 16 *	1,9,10	G06F17/30
A	US 5 708 828 A (COLEMAN J TODD) 13. Januar 1998 (1998-01-13) * Zusammenfassung * * Ansprüche *	1,9,10	
A	US 5 809 317 A (KOGAN DAN D ET AL) 15. September 1998 (1998-09-15) * Zusammenfassung * * Spalte 5, Zeile 3 - Spalte 5, Zeile 40 *	1,10	
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt			RECHERCHIERTE SACHGEBIETE  G06F
Recherchen <b>DEN HAAG</b>		Abschlußdatum der Recherche <b>3. August 1999</b>	Prüfer <b>Abbing, R</b>
KATEGORIE DER GEZEICHNETEN DOKUMENTE			
X: von besonderer Bedeutung allein betrachtet Y: von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A: technischer Hintergrund P: nichttechnische Offenbarung Z: Zwischenliteratur		T: der Erfindung zugrunde liegende Theorien oder Grundsätze E: dieses Patendokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist O: in der Anmeldung angeführtes Dokument C: aus anderen Gründen angeführtes Dokument A: Mitglied der gleichen Patentfamilie, überreichtendes Dokument	

EPO FORM 1533 (02/97) (PAC/2)

**ANHANG ZUM EUROPÄISCHEN RECHERCHENBERICHT  
ÜBER DIE EUROPÄISCHE PATENTANMELDUNG NR.**

EP 99 10 3194

In diesem Anhang sind die Mitglieder der Patentfamilien der im obengenannten europäischen Recherchenbericht angeführten Patentdokumente angegeben.  
Die Angaben über die Familienmitglieder entsprechen dem Stand der Daten des Europäischen Patentamts am  
Diese Angaben dienen nur zur Unterrichtung und erfolgen ohne Gewähr.

03-08-1999

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglieder der Patentfamilie		Datum der Veröffentlichung
US 5649192	A	15-07-1997	JP	7006077 A	10-01-1995
US 5708828	A	13-01-1998	AU	5873796 A	11-12-1996
			WO	9637817 A	28-11-1996
US 5809317	A	15-09-1998	KEINE		

EP 1 030 254 A1

Für nähere Einzelheiten zu diesem Anhang : siehe Amtsblatt des Europäischen Patentamts, Nr. 12/82